THE JOURNAL OF PHYSICAL CHEMISTRY Letters

pubs.acs.org/JPCL

Letter

# Dynamic Asymmetry Exposes 2019-nCoV Prefusion Spike

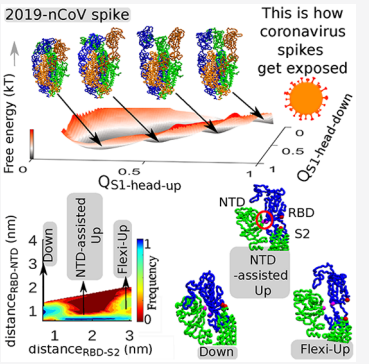
Susmita Roy,* Akhilesh Jaiswar, and Raju Sarkar

Cite This: *J. Phys. Chem. Lett.* 2020, 11, 7021−7027

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

**ABSTRACT:** The novel coronavirus (2019-nCoV) spike protein is a smart molecular machine that instigates the entry of coronavirus to the host cell causing the COVID-19 pandemic. In this study, a symmetry-information-loaded structure-based Hamiltonian is developed using recent Cryo-EM structural data to explore the complete conformational energy landscape of the full-length prefusion spike protein. The study finds the 2019-nCoV prefusion spike to adopt a unique strategy by undertaking a dynamic conformational asymmetry that results in two prevalent asymmetric structures of spike where one or two spike heads rotate up to provide better exposure to the host-cell receptor. A few unique interchain interactions are identified at the interface of closely associated N-terminal domain (NTD) and receptor binding domain (RBD) playing a crucial role in the thermodynamic stabilization of the up conformation of the RBD in the case of the 2019-nCoV spike. The interaction-level information decoded in this study may provide deep insight into developing effective therapeutic targets.




We are in the midst of a global catastrophic situation due to coronavirus outbreak where every day the global death toll is beating its past count. While history has witnessed past pandemics,[1−3] 2019 Novel Coronavirus (2019-nCoV) trends to outcompete them all by its rapid transmission in a short period to win the crown. As the name "corona" (in Latin, it means crown), the accused of the outbreak makes use of a "crown-shaped" molecular machine, the trimeric spike protein to drive the virus entry into the host cells. 2019-nCoV is the newest addition of betacoronavirus genus.[4] While the sequence and structural similarity to the severe acute respiratory syndrome coronavirus spike (SARS-CoV S) repute the identity of 2019-nCoV spike as SARS-CoV-2 S,[5,6] a recent study reported that the SARS-CoV-2 S has 10−20 fold higher affinity to human angiotensin-converting enzyme 2 (ACE2) receptor than that of SARS-CoV S.[7]

The large ectodomain of the S glycoprotein of the coronavirus uses the S1 subunit for receptor binding and the trimeric S2 stalk for host-cell membrane fusion (Figure 1A).[8,9] In SAR-CoV-2 S glycoprotein, the $\beta$-sheet-rich S1 subunit comprises an N terminal domain (NTD) and a receptor-binding domain (RBD) toward its C-terminal domain (CTD) (Figure 1B). So far, static structural characterization reported that the RBD of S1 has an intrinsic hinge-like conformational movement that generates the "up" and "down" conformations.[7,8,10] Other betacoronaviruses, like SARS-CoV, MERS-CoV, and distantly related alphacoronavirus porcine epidemic diarrhea virus (PEDV) also have this apparently stochastic RBD movement.[11,12] The combination of RBD up−down rearrangement may lead each S1-head of the trimeric prefusion spike protein of coronavirus to adopt different possible conformations: (i) 3down, (ii) 1up−2down, (iii) 2up−1down, and (iv) 3up (Figure 1C). Among them 3down, 3up

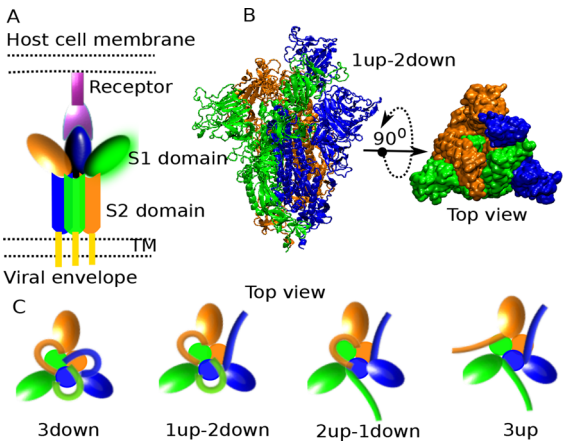



**Figure 1.** Conformational illustration of coronavirus spike protein. (A) Schematic of receptor-bound spike protein including the receptor-binding subunit, S1 and the membrane-fusion subunit, S2 of coronavirus are shown. (B) Side and top views of the homotrimeric structure of SARS-CoV-2 spike protein with one RBD of the S1 subunit head rotated in the up conformation. This all-atomistic conformation is taken from the pdb id: 6vsb. (C) RBD up−down movement is expected to lead S1 heads of the trimeric spike protein to attain the following possible conformers: (i) 3down (ii) 1up−2down (iii) 2up−1down, and (iv) 3up. This is an analogue demonstration of the spike protein top-view where NTDs are represented by colored ovals, RBDs are represented by flexible sticks and S2 domains are represented by filled circles.

**Received:** May 10, 2020
**Accepted:** August 3, 2020
**Published:** August 3, 2020




https://dx.doi.org/10.1021/acs.jpclett.0c01431
*J. Phys. Chem. Lett.* 2020, 11, 7021−7027

are symmetric conformers and 1up−2down, 2up−1down are asymmetric conformers. Single-particle cryo-electron microscopy (Cryo-EM) determined few such symmetric and asymmetric structures referred to as the receptor-binding inactive states and receptor-binding active states, respectively.[8] The asymmetric structure where one of the RBDs rotates up was thought to be less stable for SARS-CoV S.[10] In comparison, a recent Cryo-EM study found three RBDs in 1up−2down conformation as a predominant arrangement in the prefusion state of the 2019-nCoV S trimer.[7] This arrangement apparently appears legitimate for SARS-CoV-2 S in order to explain the higher affinity of 1up−2down for ACE2 receptor than that of SARS-CoV S according to recent Cryo-EM data.[7] However, we cannot rule out the possibility of 2up−1down conformation also as a functional state, which may provide even stronger binding with ACE2 considering the fact that ACE2 is a dimeric receptor.[9,13] This is a mere hypothesis, which certainly needs experimental validation. Yet, this hypothesis is consistent with a recent crystallographic study demonstrating that CR3022, a neutralizing antibody isolated from convalescent SARS patients targets the RBD when at least two RBD units on the trimeric spike protein are in the up conformation.[14] Assembling all these experimental results, it is high time to understand the molecular mechanism of S1-head coordination of trimeric SAR-CoV-2 S and to identify important interaction in regulating spike up−down conformations.

A major challenge of this study was simulating the gigantic structure of the full-length trimeric spike, as it is associated with the large scale conformational transition. It is indeed a daunting task to explore the full conformational landscape at an atomic length scale. To overcome this, a structure-based coarse-grained molecular dynamic simulation approach has been adopted.[15] The simulation started with a full-length homotrimeric spike protein structure generated from homology modeling that involves the alignment of a target sequence and a template structure (pdb: 6vsb).[7,16] This also helped to build the missing loops. The domain-specific residue range for the full-length, trimeric SARS-CoV-2 S is given in Figure 2A.

The S1 head movement of the trimeric spike is captured by developing a supersymmetric topology-based modeling framework (Figure 2B) (described in the Method pipeline in the Supporting Information). With this, the molecular machine is ready to swing each of its S1 heads between its "up" and "down" conformations (Movie S1). A number of Cryo-EM structures captured the "up" and "down" conformations of the RBD domain of spike proteins of other coronaviruses including SARS-CoV-2 where the S1 subunit undergoes a hinge-like conformational movement prerequisite for receptor binding (Figure 2C).[7,8,10,17] Apart from the hinge-responsive RBD−cleft interactions, in this study, a few interchain interactions are found to assist the "RBD-up" and the "RBD-down" conformations (shown in Figure 2D,E, Movie S2 in the Supporting Information). These interactions are identified to impact the breathing of RBD of SARS-CoV-2 S. This makes the early referred "RBD-up/down" conformations slightly different from the "S1-head-up/down" conformation for the trimeric SARS-CoV-2 S, as clearly the former is regulated only by intrachain interactions while the latter is regulated by both intra- and interchain interactions (Figure S1 in the Supporting Information). After identifying all these unique intra- and interchain contacts[18,19] extracted from the corresponding "S1-head-up" and "S1-head-down" conformations, a supersymmetric contact map is generated. This follows the development of a structure-based model Hamiltonian (Methods in the Supporting Information), which is based on the energy landscape theory of protein folding.[20−24] This approach potentiates the trimeric spike not only to adopt C3 symmetric "3up" and "3down" states but also to break the symmetry in a thermodynamically governed way (Figures S1−S4 in the Supporting Information).[25,26]

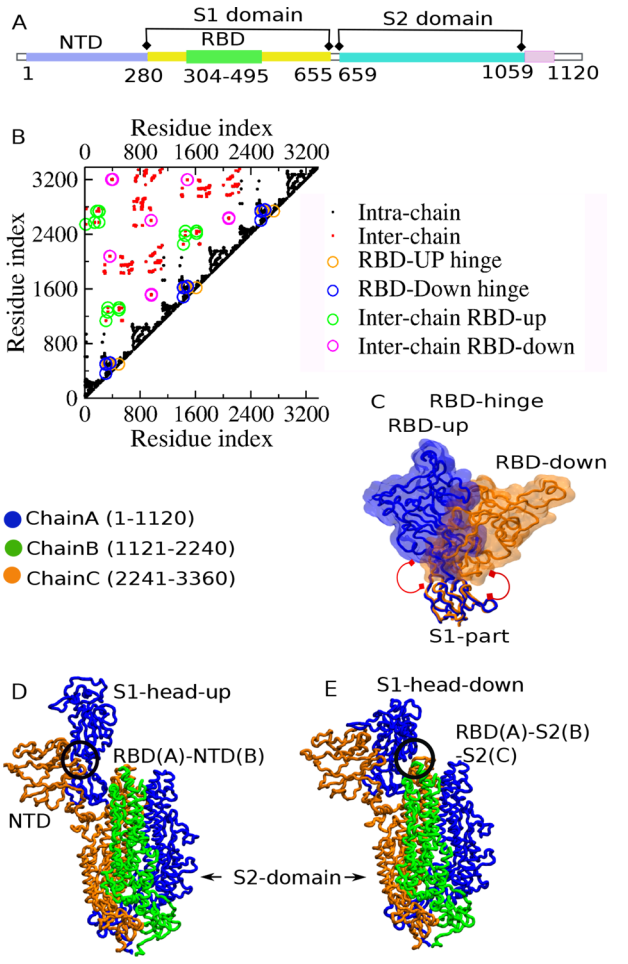



**Figure 2.** Building a supersymmetric contact map of the homotrimeric SARS-CoV-2 spike protein. (A) Amino acid sequence ranges of NTD, RBD, and S2-subunit are only highlighted. (B) Residue−residue native contact map identifying unique intra- and interchain contact-pairs formed by any single monomer in its S1-head up and S1-head down states. (C) Within intrachain contacts, the unique contacts that drive hinge motion leading to RBD-up and RBD-down states are highlighted in the structure, as well as in the contact map. (D) Interchain unique contacts between RBD and NTD domains upholding the S1-head-up state. (E) Interchain unique contacts are responsible for connecting the RBD of ChainA with the S2-stalk of ChainB and the S2 stalk of ChainC.

To monitor the transition between the "S1-head-up" and the "S1-head-down" states for each monomer with the trimeric interactions, a large pool of unbiased long time trajectories were generated where multiple occurrences of up and down states for each monomer have been sampled. We employ a reaction coordinate, $Q$, the fraction of the native contact[19,27] corresponding to the interchain contacts associated with the "S1-head-up" and the "S1-head-down" states. A typical trajectory plot of $Q$ extracted from the equilibrium simulation of the trimeric prefusion spike clearly shows the hopping


https://dx.doi.org/10.1021/acs.jpclett.0c01431
J. Phys. Chem. Lett. 2020, 11, 7021−7027

between different conformational states as hypothesized earlier (Figure 3A). Furthermore, the dynamic transitions between

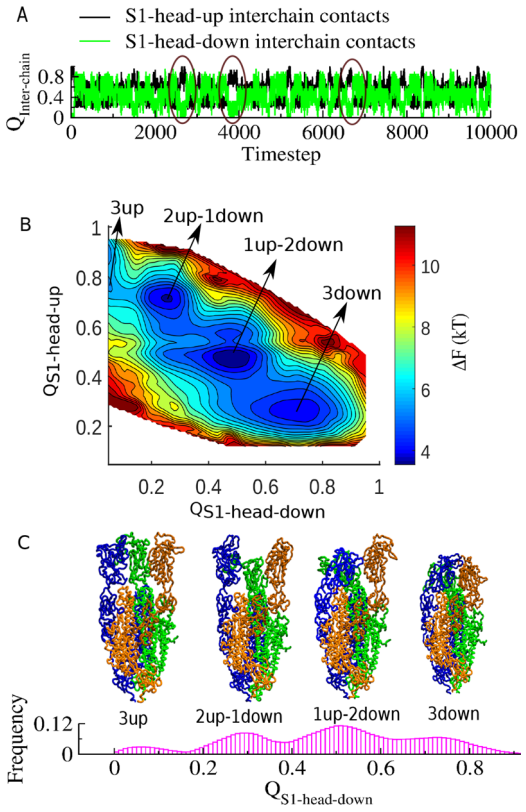



**Figure 3.** Conformational transition of SARS-CoV-2 spike protein in its prefused state. (A) Fraction of native contact ($Q$) dynamics counting interchains contact pairs formed in the S1-head-up state and the S1-head-down state. (B) A two-dimensional free energy landscape of the conformational transition as a function of interchain contacts supporting S1-head-down ($x$-axis) and S1-head-up state ($y$-axis) explores all possible conformations. (C) The representative structure corresponding to each minimum of the free energy landscape is designated as follows: (i) 3up, (ii) 2up−1down, (iii) 1up−2down, and (iv) 3down state (as shown in the one-dimension population distribution plot).

those two major asymmetric states (1up−2down, $Q_{S1\text{-head-down}} \approx 0.45$; 2up−1down, $Q_{S1\text{-head-down}} \approx 0.15$) are evident in the representative $Q$-trajectory. Analysis of all the simulations yields a 2-D free energy landscape of the trimeric spike protein of SARS-CoV-2 (Figure 3B) with its all possible conformations. The conformations corresponding to the minima of the free energy landscape are shown in Figure 3C. The temperature dependence of conformational transition indicates that the configurational entropy and enthalpy compensation results in the enhanced population of the asymmetric 1up−2down to 2up−1down conformations (Figure S4 in the Supporting Information). While the predominant population of the 1up−2down state is consistent with the recent Cryo-EM data[7] (Movie S1 in the Supporting Information), the other asymmetric structure (2up−1down) emerges as a best binding epitope for CR3022 (an antibody collected from convalescent SARS patients) according to a recent antibody recognition study of SARS-CoV-2 S.[14]

From Figure 2 and Figure 3, we have identified crucial inter- and intrachain interaction sites that essentially control the conformational dynamics of SARS-CoV-2 S. This elevated our curiosity to further analyze and compare interchain interaction sites similar to those present in SARS-CoV S and MERS-CoV S. In this study, sequence and interaction level (Figure 4 and Figures S5−S7 in the Supporting Information) comparisons

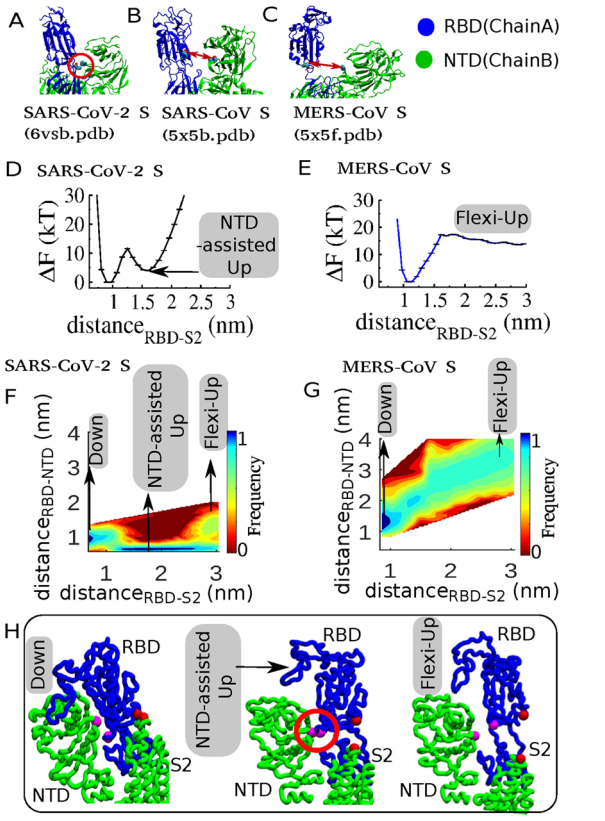



**Figure 4.** Free energy calculations of up−down transition demonstrate that the interchain RBD−NTD interaction stabilizes the up conformation of SARS-CoV-2 S. (A) Unique interchain interactions formed by the RBD of one chain with the NTD of the adjacent chain stabilizing the S1-head-up conformation in SARS-CoV-2 S (pdb: 6vsb). The interchain domain closure is analyzed by the interchain proline−proline (residue id, P230−P521; distance, 0.65 nm) distance measurement. The same distance measured for the following spikes: (B) SARS-CoV spike (pdb, 5x5b; residue id, P223−P324; distance, 2.77 nm) and (C) MERS-CoV spike (pdb, 5x5f; residue id, P285−P586; distance, 3.13 nm). (D) Average free energy profile of the S1-head up−down transition of a single chain of the SARS-CoV-2 S trimer. Note the NTD-assisted up-state stabilization. (E) Average free energy profile of the S1-head up−down transition of a single chain for MERS-CoV S trimer. Note the broad ensemble of flexible up-state (flexi-up). (F) 2D-landscape pathway of the up−down transition as a function of RBD−NTD and RBD−S2 stalk distances for the SARS-CoV-2 S trimer. The up−down pathway is intervened by the "NTD-assisted up" state. (G) 2D-landscape pathway of the up−down transition as a function of RBD−NTD and RBD−S2 stalk distances for MERS-CoV S trimer. (H) Representative structure corresponding to each minimum associated with the free energy profile of the up−down transition for SARS-CoV-2 S. The free energy order parameter is the interchain RBD−S2 subunit distance, which is characterized by the distance between a lysine (residue id: 360) of the RBD and an aspartate of the adjacent S2-stalk (residue id: 2079) (potential salt-bridge interaction, marked in red). To characterize the pathway in 2D, along the free energy coordinate the interchain RBD−NTD closure is monitored by the distance between a proline (residue id: 495) of the RBD and another proline (residue id: 1324) of the adjacent NTD (marked in magenta).



https://dx.doi.org/10.1021/acs.jpclett.0c01431
J. Phys. Chem. Lett. 2020, 11, 7021−7027

have been made over the Cryo-EM structures of SARS-CoV-2 S (pdb: 6vsb), SARS-CoV S (pdb: 5x5b), and MERS-CoV S (pdb: 5x5f).[7,12] This comparison elucidates that in the case of SARS-CoV-2 S the NTD and RBD are closely interacting. Among those close interactions, the key interaction involves a proline residue of the RBD forming CH-$\pi$ type[28] interaction with a tyrosine residue of the adjacent NTD. The same proline is also involved in a hydrophobic interaction with another proline and isoleucine residing on the NTD of the adjacent chain (Figure 4A, Figure S1 in the Supporting Information). Interchain proline−proline distance measurements show that the corresponding RBD−NTD domains are far away in the case of SARS-CoV S (Figure 4B) (residue id, P223−P324 as per pdb id 5x5b; distance, 2.77 nm) and further away in the case of MERS-CoV S (Figure 4C) (residue id, P285−P586 as per pdb id 5x5f and P268-P2748 as per our model index; distance, 3.13 nm) as compared to that of SARS-CoV-2 S (residue id, P230−P521 as per pdb id 6vsb and P495-P1324 as per our model index; distance, 0.65 nm). These measurements from their respective Cryo-EM structures clearly demonstrate that while in SARS-CoV-2 S, the interchain RBD−NTD units are closely interacting, in the case of SARS or MERS-CoV S, they are not. To further understand the effect of interchain RBD−NTD domain closure on the S1-head up−down movement, we have calculated the average free energy of the S1-head up−down transition of any given chain where S1 domains of the other two chains are restrained to remain in the down conformation. We have adopted the umbrella sampling method to calculate the average free energy profile as a function of the distance between the RBD domain and the S2 stalk.[29] The average free energy profile of the S1-head up−down transition of SARS-CoV-2 S has been compared with that of MERS-CoV S, which holds a representative structure where interchain RBD and NTD are far apart from each other. The structural characterization for the MERS-CoV S trimer follows the same approach as we have described for SARS-CoV-2 S using our method pipeline (Methods in the Supporting Information). This involves the identification of the unique intrachain and interchain contacts[18,19] extracted from its corresponding "S1-head-up" and "S1-head-down" conformations, generating a supersymmetric contact map and the development of a structure-based model Hamiltonian for MERS-CoV S (Figure S6 in the Supporting Information). Although SARS-CoV-2 S and MERS-CoV S are two different protein trimers of very different system sizes, it is quite obvious that their folding temperature and the temperature dependence of free energy would be different. Besides, comparing two free energy profiles we find that the free energy barrier between the S1-head-up state and the S1-head-down state for MERS-CoV S near its up−down transition temperature is significantly large. Due to this large free energy barrier, we could not sample all possible up−down conformations for MERS-CoV S by canonical equilibrium simulations as we obtained for SARS-CoV-2 S in Figure 3C. Instead, extensive free energy calculations as a function of RBD−S2 stalk distance helped us to capture the up−down transition for MERS-CoV S and compare the transition mechanism and pathway between SARS-CoV-2 S and MERS-CoV S. The free energy profile for SARS-CoV-2 S reveals two distinct minima (Figure 4D). They involve a down state and a less flexible up state. In contrast, for MERS-CoV S the free energy profile reveals a distinct down-state minimum along with an extended, flat basin, which involves a broad ensemble of flexible up conformations that we call a "flexi-up" conformation (Figure 4E and Figure S7). To monitor how the RBD−NTD connection evolves during this transition, we analyzed the characteristic distance (shown in Figure 4A,C) between the RBD and the adjacent NTD along the free energy coordinate which illuminates the pathway of the up−down transition. The up−down transition for SARS-CoV-2 S in the presence of RBD−NTD interaction follows a distinct ordered pathway via stable "NTD-assisted-up" conformations (Figure 4F). For MERS-CoV S, the transition pathways appear less ordered/scattered due to the high degree of flexibility of its up conformations (Figure 4G). The representative structure of (i) down, (ii) NTD-assisted-up, and (iii) flexi-up conformations of SARS-CoV-2 S are shown in Figure 4H. The representative structure corresponding to each basin on the energy landscape serves as a landmark of the NTD guided ordered and nonguided scattered pathways. The representative conformations for MERS-CoV S are shown in Figure S7. The high degrees of flexibility of the up conformations and a similar scattered pathway (as we obtained for MERS-CoV S) have also been obtained in the case of SARS-CoV-2 S when the relevant NTD−RBD contacts are deleted from its model Hamiltonian (Figures S8 and S9 in the Supporting Information). These important contacts are enlisted in Table S2. The interchain RBD−NTD connection, in the case of SARS-CoV-2 S, is also found to impact the RBD−hinge interaction by upregulating more RBD-up conformation. In the absence of such interchain interactions, the RBD mostly stays in the down conformation, allowing the RBD to break the symmetry rarely in a stochastic manner (Figure S10). Such stochastic RBD movement has also been observed in early studies for other closely related betacoronaviruses SARS-CoV and MERS-CoV, and the more distantly related alphacoronavirus porcine epidemic diarrhea virus.[11,12] The RBD−NTD connections render a significant difference in the free energy landscapes of SARS-CoV-2 S and MERS-CoV S by inducing the stabilization of the up conformation of the S1-domain of SARS-CoV-2 S. This comparison encouraged us to perform an in-silico mutagenic analysis focusing on the interactions present at the interface of interchain RBD and NTD, in the case of SARS-CoV-2 S. There are limited methods to study the effects of mutation on protein dynamics. *DynaMut*, a Web server interface that uses well-established normal modes analysis approach to study the mutation effects on the protein dynamics at an atomistic level (Method, Supporting Information).[30,31] Based on the calculated Gibb's free energy/folding free energy, DynaMut helps us to analyze which site- mutation causes significant destabilization effects on the structure of SARS-CoV-2 S trimer. The potential sites/hotspots are listed in Table S3 and Table S4. We have also analyzed the effects of site- mutation by analyzing residue-specific structural flexibility [rms fluctuation (rmsf)] of the trimer. These sites are highlighted in Figure S11. Mutagenic analysis suggests that the aforesaid proline including nearby histidine (the same histidine that we find in our sequence alignment analysis in Figure S5) (on RBD) mediated interactions may have important effects on the stability of interchain RBD−NTD association. Apart from this interchain RBD−NTD association, the assistance from the interchain RBD−S2-stalk related interfacial contacts are also found to modulate the population dynamics of the RBD-down conformation (Figure S12 in the Supporting Information). The influence of this interchain RBD−S2-stalk interaction has also been observed in early Cryo-EM analyses where two


https://dx.doi.org/10.1021/acs.jpclett.0c01431
J. Phys. Chem. Lett. 2020, 11, 7021−7027

proline mutations at the top of the S2 stalk (inferring RBD−S2 interchain connection) helped to stabilize the "up" conformers of SARS-CoV S.[32]

Below, we summarize the main features of our study:

(i) A supersymmetric contact map is generated for the full-length trimeric SARS-CoV-2 spike protein, which successfully isolates all the unique intra- and interchain interactions. This supersymmetric contact-map generation approach can be implemented to any other trimeric spikes that are amenable to adopt both the symmetric and asymmetric conformations in a thermodynamically governed way. In most betacoronaviruses, each chain of the trimer is naturally programmed to swing up and down from its S2-stalk. This study not only distinguishes the unique intrachain interactions that maintain RBD hinge motion but also identifies the crucial interchain interactions between RBD−NTD (if exist) and RBD−S2 stalk related domain−domain interactions regulating the stability of S1-head conformations for SARS-CoV-2 S and MERS-CoV S.

(ii) The symmetry-information-loaded structure-based Hamiltonian developed in this study helps to explore the full conformational landscape of SARS-CoV-2 S. The study finds a dynamic transition between two prevalent asymmetric structures of the spike near their up−down transition temperature. As the temperature goes down, the population shifts more toward the down conformation.

(iii) The most interesting finding of this study is primarily based on a comparison among the resolved Cryo-EM structures of SARS-CoV-2 S, SARS-CoV S, and MERS-CoV S. This study identifies that while in SARS-CoV-2 S interchain RBD−NTD are closely interacting, in SARS-CoV S and MERS-CoV S, they are not.

(iv) We have adopted the umbrella sampling free energy simulation technique to capture the S1-head up−down transition. We have compared the transition pathway of SARS-CoV-2 S where RBD−NTD association exists with the same of the representative betacoronavirus and MERS-CoV S where RBD−NTD are distant from one another. The study finds the interchain RBD−NTD closure significantly stabilizes the S1-head-up state in the case of SARS-CoV-2 S. The study also characterizes the scattered pathway of the up−down transition for MERS-CoV S and the modified version of SARS-CoV-2 S where NTD−RBD interactions are missing.

(v) By mutagenic analysis, the study identifies a few key residues in the interface regime of interchain RBD and NTD that have the potential to affect the dynamics and thermodynamics of interchain RBD−NTD interaction, in turn, the stabilization of the S1-head-up state.

The synergy between intrachain RBD−hinge interactions and interchain interactions allows the trimeric SARS-CoV-2 S to adopt a dynamical feature unique from other corona-virus spikes like MERS-CoV S. It appears that the interchain interactions driven rapid symmetry breaking strategy potentiates this spike machine to turn on its receptor-binding propensity by activating the stable "up" state conformation. A number of early studies along with the recent Cryo-EM structure of full-length human ACE2 confirm that an up conformation of RBD is required to bind to the receptor.[7,9,10] The binding of the RBD in its down state to ACE2 faces steric hindrance, suggesting that the "up conformation" represents a receptor-binding competent state.[10] This naturally poses the question of which interactions stabilize the formation of a stable up conformation. Although the study finds the importance of this interchain NTD−RBD association in stabilizing the up conformation, this is beyond the scope of this study to explain whether these NTD−RBD interactions have the potential to tune the binding affinity of the "up" conformation to the receptor, ACE2. This certainly warrants future experiments.

During the dynamical evolution process of a structure, the possibility of the formation of several non-native interactions adds more complexity to this problem. Those were excluded adopting this minimalist modeling approach ensuring that the native interactions contain sufficient information describing the interactions present in this biomolecular system.[15,21,22,25,33−38] This approach also reduces the complexity arises from different force fields. It is worth mentioning here that in this particular work the C-alpha structure-based modeling approach is adopted, solely keeping in mind that here our purpose is to capture the large conformational changes of such a system, which is of huge length scale (∼3500 amino acid). However, this reduced model has its own limitations where we may miss microscopic events associated with possible non-native interactions, side chain interactions, energetic heterogeneity and explicit electrostatic and solvent mediated interactions.[39−42]

The dynamic asymmetry induced by the identified unique interchain interactions in this study highlights a different mechanism for SARS-CoV-2 S stabilizing more up conformation. During the conformational transition between two prevalent asymmetric structures, where one or two S1-heads among three heads are rotated up, most likely offer stable exposure to the host-cell receptor. The asymmetric structure where one of the three S1-head rotated up is consistent with the recent Cryo-EM structure of the 2019-nCoV prefusion spike.[7] The other asymmetric structure where two of the three S1-heads rotated up emerges as a best binding epitope for CR3022 (an antibody collected from a convalescent SARS patient) according to a recent antibody recognition study of SARS-CoV-2 S.[14]

Although in the current situation developing diagnostics and antiviral therapies is of utmost priority, we believe the present structure-based-model-derived information at the microscopic interaction level might provide deep insight into design effective decoys or vaccine to fight 2019-nCoV infection.

## ■ ASSOCIATED CONTENT

### Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jpclett.0c01431.

Method pipeline of building super-symmetric contact map of SARS-CoV-2 prefusion spike protein, detailed methodology for super-symmetric contact map generation, development of the structure-based Hamiltonian of trimeric spike protein simulation, equilibrium simulation details, temperature-dependent simulations and analyses, general free energy calculation method, umbrella sampling method for free energy calculations, mutagenic analysis method, figures of the inter-chain interaction, structural alignment of two chains, RMS deviation of each chain, temperature dependence of



https://dx.doi.org/10.1021/acs.jpclett.0c01431
J. Phys. Chem. Lett. 2020, 11, 7021−7027

transitions, sequence alignments, structure, contact map, 2D landscape pathways, free energy profiles, RBD hinge dynamics, and mutagenic analysis images; tables of parameter values, inter-chain NTD−RBD interfacial contact information for SARS-CoV-2 S, and mutagenic analysis data; references (PDF)

Movie of conformational dynamics of full-length trimeric SARS-CoV-2 spike protein showing rapid symmetry breaking (MP4)

Movie of conformational dynamics of a monomer of the full-length SARS-CoV-2 S showing RBD hinge motion (MP4)

## AUTHOR INFORMATION


### Corresponding Author

**Susmita Roy** − *Department of Chemical Sciences, Indian Institute of Science Education and Research Kolkata, Mohanpur, West Bengal 741246, India;* orcid.org/0000-0001-6411-4347; Email: susmita.roy@iiserkol.ac.in

### Authors

**Akhilesh Jaiswar** − *Department of Chemical Sciences, Indian Institute of Science Education and Research Kolkata, Mohanpur, West Bengal 741246, India*

**Raju Sarkar** − *Department of Chemical Sciences, Indian Institute of Science Education and Research Kolkata, Mohanpur, West Bengal 741246, India*

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jpclett.0c01431


### Notes

The authors declare no competing financial interest.

All data and codes used in the analysis are available from the corresponding author to any researcher for purposes of reproducing or extending the analysis under a material transfer agreement with IISER-Kolkata, India.

## ACKNOWLEDGMENTS

The authors thank the DIRAC supercomputing facility at IISER-Kolkata for computational support. S.R. acknowledges support from the Department of Biotechnology (DBT), India (Grant No. BT/12/IYBA/2019/12).

## REFERENCES


(1) *Microbial Evolution and Co-Adaptation: A Tribute to the Life and Scientific Legacies of Joshua Lederberg: Workshop Summary*; National Academic Press: Washington, DC, 2009.

(2) McNeill, W. H. *Plagues and peoples*; Anchor Books: New York, 1998; 365 pp.

(3) Zheng, J. SARS-CoV-2: an Emerging Coronavirus that Causes a Global Threat. *Int. J. Biol. Sci.* **2020**, *16* (10), 1678−1685.

(4) Xu, X.; Chen, P.; Wang, J.; Feng, J.; Zhou, H.; Li, X.; Zhong, W.; Hao, P. Evolution of the novel coronavirus from the ongoing Wuhan outbreak and modeling of its spike protein for risk of human transmission. *Sci. China: Life Sci.* **2020**, *63* (3), 457−460.

(5) Lam, T. T.; Shum, M. H.; Zhu, H. C.; Tong, Y. G.; Ni, X. B.; Liao, Y. S.; Wei, W.; Cheung, W. Y.; Li, W. J.; Li, L. F. Identifying SARS-CoV-2 related coronaviruses in Malayan pangolins. *Nature* **2020**, *583*, 282.

(6) Li, F. Structure, Function, and Evolution of Coronavirus Spike Proteins. *Annu. Rev. Virol.* **2016**, *3* (1), 237−261.

(7) Wrapp, D.; Wang, N.; Corbett, K. S.; Goldsmith, J. A.; Hsieh, C. L.; Abiona, O.; Graham, B. S.; McLellan, J. S. Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. *Science* **2020**, *367* (6483), 1260−1263.

(8) Walls, A. C.; Park, Y. J.; Tortorici, M. A.; Wall, A.; McGuire, A. T.; Veesler, D. Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein. *Cell* **2020**, *181*, 281.

(9) Yan, R.; Zhang, Y.; Li, Y.; Xia, L.; Guo, Y.; Zhou, Q. Structural basis for the recognition of SARS-CoV-2 by full-length human ACE2. *Science* **2020**, *367* (6485), 1444−1448.

(10) Gui, M.; Song, W.; Zhou, H.; Xu, J.; Chen, S.; Xiang, Y.; Wang, X. Cryo-electron microscopy structures of the SARS-CoV spike glycoprotein reveal a prerequisite conformational state for receptor binding. *Cell Res.* **2017**, *27* (1), 119−129.

(11) Wrapp, D.; McLellan, J. S. The 3.1-Angstrom Cryo-electron Microscopy Structure of the Porcine Epidemic Diarrhea Virus Spike Protein in the Prefusion Conformation. *J. Virol.* **2019**, *93* (23), e00923-19.

(12) Yuan, Y.; Cao, D.; Zhang, Y.; Ma, J.; Qi, J.; Wang, Q.; Lu, G.; Wu, Y.; Yan, J.; Shi, Y.; et al. Cryo-EM structures of MERS-CoV and SARS-CoV spike glycoproteins reveal the dynamic receptor binding domains. *Nat. Commun.* **2017**, *8*, 15092.

(13) Hoffmann, M.; Kleine-Weber, H.; Schroeder, S.; Kruger, N.; Herrler, T.; Erichsen, S.; Schiergens, T. S.; Herrler, G.; Wu, N. H.; Nitsche, A. SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor. *Cell* **2020**, *181*, 271.

(14) Yuan, M.; Wu, N. C.; Zhu, X.; Lee, C. D.; So, R. T. Y.; Lv, H.; Mok, C. K. P.; Wilson, I. A. A highly conserved cryptic epitope in the receptor-binding domains of SARS-CoV-2 and SARS-CoV. *Science* **2020**, *368*, 630.

(15) Clementi, C.; Nymeyer, H.; Onuchic, J. N. Topological and energetic factors: what determines the structural details of the transition state ensemble and "en-route" intermediates for protein folding? An investigation for small globular proteins. *J. Mol. Biol.* **2000**, *298* (5), 937−53.

(16) Waterhouse, A.; Bertoni, M.; Bienert, S.; Studer, G.; Tauriello, G.; Gumienny, R.; Heer, F. T.; de Beer, T. A. P.; Rempfer, C.; Bordoli, L.; et al. SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res.* **2018**, *46* (W1), W296−W303.

(17) Li, Z.; Tomlinson, A. C.; Wong, A. H.; Zhou, D.; Desforges, M.; Talbot, P. J.; Benlekbir, S.; Rubinstein, J. L.; Rini, J. M. The human coronavirus HCoV-229E S-protein structure and receptor binding. *eLife* **2019**, *8*, e51230.

(18) Noel, J. K.; Levi, M.; Raghunathan, M.; Lammert, H.; Hayes, R. L.; Onuchic, J. N.; Whitford, P. C. SMOG 2: A Versatile Software Package for Generating Structure-Based Models. *PLoS Comput. Biol.* **2016**, *12* (3), No. e1004794.

(19) Noel, J. K.; Whitford, P. C.; Onuchic, J. N. The shadow map: a general contact definition for capturing the dynamics of biomolecular folding and function. *J. Phys. Chem. B* **2012**, *116* (29), 8692−702.

(20) Leopold, P. E.; Montal, M.; Onuchic, J. N. Protein folding funnels: a kinetic approach to the sequence-structure relationship. *Proc. Natl. Acad. Sci. U. S. A.* **1992**, *89* (18), 8721−5.

(21) Wolynes, P. G. Symmetry and the energy landscapes of biomolecules. *Proc. Natl. Acad. Sci. U. S. A.* **1996**, *93* (25), 14249−55.

(22) Wolynes, P. G.; Eaton, W. A.; Fersht, A. R. Chemical physics of protein folding. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109* (44), 17770−1.

(23) Zwanzig, R.; Szabo, A.; Bagchi, B. Levinthal's paradox. *Proc. Natl. Acad. Sci. U. S. A.* **1992**, *89* (1), 20−2.

(24) Bryngelson, J. D.; Onuchic, J. N.; Socci, N. D.; Wolynes, P. G. Funnels, pathways, and the energy landscape of protein folding: a synthesis. *Proteins: Struct., Funct., Genet.* **1995**, *21* (3), 167−95.

(25) Jana, B.; Hyeon, C.; Onuchic, J. N. The origin of minus-end directionality and mechanochemistry of Ncd motors. *PLoS Comput. Biol.* **2012**, *8* (11), No. e1002783.

(26) Lin, X.; Eddy, N. R.; Noel, J. K.; Whitford, P. C.; Wang, Q.; Ma, J.; Onuchic, J. N. Order and disorder control the functional

rearrangement of influenza hemagglutinin. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111* (33), 12049−54.
(27) Nymeyer, H.; Socci, N. D.; Onuchic, J. N. Landscape approaches for determining the ensemble of folding transition states: success and failure hinge on the degree of frustration. *Proc. Natl. Acad. Sci. U. S. A.* **2000**, *97* (2), 634−9.
(28) Brandl, M.; Weiss, M. S.; Jabs, A.; Suhnel, J.; Hilgenfeld, R. C-H···pi-interactions in proteins. *J. Mol. Biol.* **2001**, *307* (1), 357−77.
(29) Torrie, G. M.; Valleau, J. P. Monte-Carlo Free-Energy Estimates Using Non-Boltzmann Sampling - Application to Subcritical Lennard-Jones Fluid. *Chem. Phys. Lett.* **1974**, *28* (4), 578−581.
(30) Rodrigues, C. H.; Pires, D. E.; Ascher, D. B. DynaMut: predicting the impact of mutations on protein conformation, flexibility and stability. *Nucleic Acids Res.* **2018**, *46* (W1), W350−W355.
(31) Rodrigues, C. H. M.; Myung, Y.; Pires, D. E. V.; Ascher, D. B. mCSM-PPI2: predicting the effects of mutations on protein-protein interactions. *Nucleic Acids Res.* **2019**, *47* (W1), W338−W344.
(32) Kirchdoerfer, R. N.; Wang, N.; Pallesen, J.; Wrapp, D.; Turner, H. L.; Cottrell, C. A.; Corbett, K. S.; Graham, B. S.; McLellan, J. S.; Ward, A. B. Stabilized coronavirus spikes are resistant to conformational changes induced by receptor recognition or proteolysis. *Sci. Rep.* **2018**, *8* (1), 15701.
(33) Brooks, C. L., 3rd; Gruebele, M.; Onuchic, J. N.; Wolynes, P. G. Chemical physics of protein folding. *Proc. Natl. Acad. Sci. U. S. A.* **1998**, *95* (19), 11037−8.
(34) Ghosh, C.; Jana, B. Intersubunit Assisted Folding of DNA Binding Domains in Dimeric Catabolite Activator Protein. *J. Phys. Chem. B* **2020**, *124* (8), 1411−1423.
(35) Honeycutt, J. D.; Thirumalai, D. The nature of folded states of globular proteins. *Biopolymers* **1992**, *32* (6), 695−709.
(36) Jana, B.; Morcos, F.; Onuchic, J. N. From structure to function: the convergence of structure based models and co-evolutionary information. *Phys. Chem. Chem. Phys.* **2014**, *16* (14), 6496−507.
(37) Jana, B.; Onuchic, J. N. Strain Mediated Adaptation Is Key for Myosin Mechanochemistry: Discovering General Rules for Motor Activity. *PLoS Comput. Biol.* **2016**, *12* (8), No. e1005035.
(38) Wolynes, P. G.; Onuchic, J. N.; Thirumalai, D. Navigating the folding routes. *Science* **1995**, *267* (5204), 1619−20.
(39) Mukherjee, A.; Bhimalapuram, P.; Bagchi, B. Orientation-dependent potential of mean force for protein folding. *J. Chem. Phys.* **2005**, *123* (1), 014901.
(40) Paci, E.; Vendruscolo, M.; Karplus, M. Native and non-native interactions along protein folding and unfolding pathways. *Proteins: Struct., Funct., Genet.* **2002**, *47* (3), 379−392.
(41) Plotkin, S. S.; Onuchic, J. N. Structural and energetic heterogeneity in protein folding. I. Theory. *J. Chem. Phys.* **2002**, *116* (12), 5263−5283.
(42) Tzul, F. O.; Schweiker, K. L.; Makhatadze, G. I. Modulation of folding energy landscape by charge-charge interactions: Linking experiments with computational modeling. *Proc. Natl. Acad. Sci. U. S. A.* **2015**, *112* (3), E259−E266.


https://dx.doi.org/10.1021/acs.jpclett.0c01431
*J. Phys. Chem. Lett.* 2020, 11, 7021−7027